RFI抑制技术在射电天文中的应用*

张海龙[1,2,3,4*]，张亚州[1,2]，王杰[1,4]，冶鑫晨[1,4]，王万琼[1]，李嘉[1]，张萌[1,2]，杜旭[1,2]

(1. 中国科学院新疆天文台，新疆 乌鲁木齐，830011；2. 中国科学院大学，北京，100049；3. 中国科学院射电天文重点实验室，江苏 南京210008；4. 国家天文科学数据中心，北京，100101)

摘 要：针对射电天文观测过程中的射频干扰(RFI)问题，详细分析了国内外台站 RFI 抑制策略。根据各天文台站实际观测过程中遇到的RFI 问题，分别从主动预防阶段、预相关阶段、后相关阶段、机器学习和深度学习等方面研究了RFI 的预防策略和抑制方法。详细分析了主动预防阶段可采取的方法，预相关阶段的自适应滤波，和空间滤波方法；后相关阶段的 VarThreshold,SumThreshold 和奇异值分解等方法。探讨了基于机器学习的主成分分析、支持向量机、全卷积神经网络、卷积神经网络、U-Net 等相关技术和方法在RFI 信号处理方面的应用。

关键词：RFI； 滤波； 阈值法； 机器学习；

中图分类号：P161 **文献标识码**： **文章编号**：

1 引言

射电天文学中，RFI 被定义为任何可能影响天文观测的有害信号，射电天文学发展过程中，抑制 RFI 一直是天文学家研究的热点。随着信息技术飞速发展，人为造成的干扰信号导致天文台站周围电磁环境恶化，影响射电望远镜正常观测。

RFI 的来源多种多样，包括外部来源和内部来源，外部来源主要由天文台址外设备产生，如人造卫星[1] (北斗和 GPS 导航卫星等)、飞机、台址附近的基站信号以及电视广播号等；内部来源主要由天文台址内使用的电子设备产生，如计算机、视频监控、网络交换设备、无线输入设备等[2]。天文信号通常为宽频带且在时间尺度上平滑变化，而 RFI 在时域和频域的幅值强度高，绝大部分 RFI 与天文信号相比有明显差异[3]。RFI 与热噪声不同，它通常由通信系统、人造雷达或电子设备产生的干扰构成，且通常具有复杂的时间和频率结构，常见通信信号的功率比天文信号高多个数量级并且会随时间变化，因此无法通过长时间积分（累积）来降低其强度，严重影响天文数据质量。天文学家针对上述问题已提出了多种 RFI 抑制方法，但仍受到以下因素制约：

* **基金项目**：国家自然科学基金(12173077, 11873082, 11803080, 12003062)；新疆维吾尔自治区天山创新团队计划(2022D14020)；中国科学院青年创新促进会；国家重点研发计划(2018YFA0404704)；国家天文科学数据中心；中国科学院天文台站设备更新及重大仪器设备运行专项经费资助；国家重点研发计划（2021YFC2203502）。

收稿日期：2022-00-00；**修订日期**：2022-00-00

作者简介：张海龙，男，正高级工程师. 研究方向:数据密集型研究. Email ：zhanghailong@xao.ac.cn

（1）近年来，随着无线技术的飞速发展及人类活动范围的扩大，地面及空间产生的人为干扰逐年递增。

（2）随着制造技术和信息技术的飞速发展，射电望远镜向大口径或大规模阵列方向发展，随着多波束或 PAF 接收系统的投入使用，使得观测设备的灵敏度及数据收集能力显著提升，同时 RFI 抑制问题更加复杂。

在射电天文观测以及数据处理过程中需要针对不同阶段采取不同的 RFI 应对策略，本文将 RFI 抑制策略分为主动预防阶段、预相关阶段及后相关阶段，不同阶段针对 RFI 问题可采取不同抑制方法。主动预防阶段若能采取有效防护手段，能够杜绝通信基站，电视广播等大部分 RFI 进入系统;预相关阶段采取参考天线及空间滤波等方法可以应对特定 RFI 源;后相关阶段中可以采取阈值法处理幅值强度远大于天文信号 RFI，或利用机器学习相关技术实现 RFI 数据标记。下面针对不同阶段可采取的 RFI 抑制策略进行详细分析:

2 RFI 主动预防阶段

2.1 电磁波宁静区

射电望远镜观测的是来自宇宙遥远天体的微弱信号，对台址周边电磁干扰有较高要求，选择地理环境较好的台址及建立无线电宁静保护区是抑制 RFI 的重要一步，可以从根源上去除大部分干扰。 500m 口径球面射电望远镜 (Five-hundred-meter Aperture Spherical radio Telescope，FAST)，是世界最大单口径、最灵敏的射电望远镜，位于贵州省黔南布依族苗族自治州平塘县克度镇大窝凼，以 FAST 台址为中心，建立半径为 30km 的 FAST 电磁波宁静区，分为不同要求的 3 个区域。其中以台址为圆心，半径 5km 范围为保护核心区，5~10km 的环带为中间区，10~30km 的环带为边远区[4]，如图 1 所示，能够屏蔽绝大部分外来干扰源。



图 1 FAST 无线电宁静保护区(来源文献[25])

Fig.1 FAST Radio Quiet Zone

新疆奇台 110 米全向可动射电望远镜(Qi Tai Radio Telescope, QTT)[5]，位于新疆昌吉州奇台县，以 QTT 台址为中心，依据 QTT 地形特征及当地实际情况，规划了 30 km 的无线电宁静区，该宁静区共分为三个区域，分别为核心区、限制区、协调区．其中核心区域为 2.5 km×4 km 的矩形区域；限制区域为 10 km×15 km 的矩形区域[6-7]，如图 2 所示。

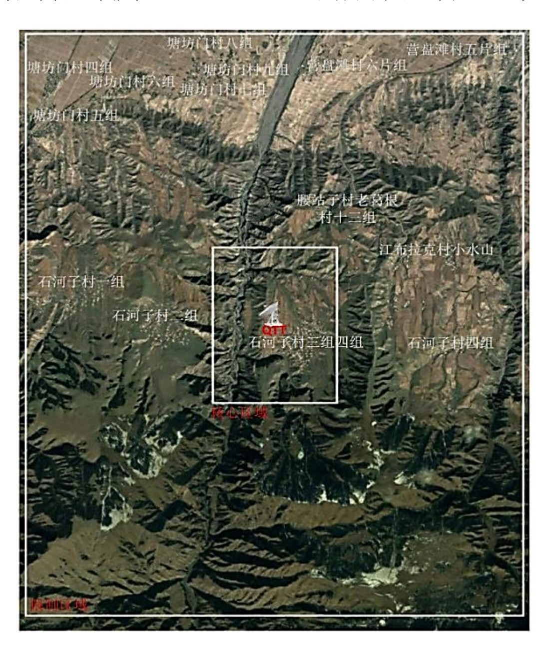


图 2 QTT 无线电宁静保护区(来源文献[6])

Fig.2 QTTRadio Quiet Zone

2.2 预留频段

国际电信联盟（ITU）通过划分频谱来协调 8.3 MHz-300 GHz 之间的无线电频谱，将

其分为几个频带，仅允许指定服务运行[1]。在分配的所有频段之间，约有 70 到 80 个频段对应于射电天文学（确切的数量取决于地区和当地法律）。针对射电天文频段，一种是射电天文专用频段（或与无源服务共享），严格禁止无线电发射；另一种是与有源业务共享的频段，只能部分施加保护。射电天文的波段选择通常与科学目标有关。例如，氢线（由于中性氢原子状态的改变而发射）位于 1420 MHz 附近，因此 1400-1427MHz 频带保留给射电天文学[2]。

即使射电天文专用频段，由于有源元件的谐波和功率泄漏，仍然可能存在干扰。对于这些情况，ITU 在其 ITU-R RA.769 建议报告中，将最大可接受的干扰定义为在测量功率中引入不超 10%的误差，这是天文学家普遍接受的阈值，该报告还为一些典型的天文波段提供了一份使用普通望远镜和观测参数的建议阈值清单[3]。

2.3 RFI 主动预防策略的不足

RFI 主动预防策略中最直接的方法是为天文台选择远离干扰源的偏远位置。自然形态（例如山脉）可以有效地阻止外部干扰，但干扰源会因为多径传播产生反射和衍射，从而增加干扰抑制难度。一些观测站研究发现，在望远镜台址周围种植松树等针叶树可有效抑制 RFI，树针中的水分可吸收 1GHz 以上的信号[4]。部分天文台站建在高海拔地区以减少大气对观测的影响且远离人类活动范围，同时尽可能避免干扰源多径传播。面对内部干扰，通常的办法是在产生干扰的电子设备（计算机、微波和射频元件）上覆盖电磁屏蔽材料，比如导电箔，也可以将电子设备集中放入屏蔽室，将干扰控制在密闭空间，并不影响设备性能。

通常，主动预防是处理干扰的最有效方法，可作为 RFI 的第一层防护。但仍然存在不足：

(1) 只能为电磁频谱的小部分提供保护。

(2) 对于高灵敏度的设备，电磁屏蔽材料并不能完全屏蔽干扰。

(3) 卫星通信以及飞机等产生的干扰，由于相对位置限制，无法通过主动预防的方法避免。

[1] http://www.fcc.gov/oet/spectrum/table/fcctable.pdf
[2] https://www.itu.int/dms_pubrec/itu-r/rec/ra/R-REC-RA.1860-0-201001-I!!PDF-E.pdf
[3] https://www.itu.int/dms_pubrec/itu-r/rec/ra/R-REC-RA.769-2-200305-I!!PDF-C.pdf
[4] https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-RA.2259-2012-PDF-E.pdf

3 RFI 预相关阶段

本文将观测获取的原始数据在未写入磁盘形成可供科学研究处理之前的 RFI 处理阶段定义为预相关阶段。对于单天线观测可利用参考天线通过自适应滤波方式去除特定干扰，对于阵列天线可采用空间滤波方法抑制 RFI。

3.1 自适应滤波

自适应滤波算法是由 Barnbaum 和 Bradley 在 1998 年首次引入射电天文领域用来解决 RFI 问题[8]，在 2005 年，Kesteven 等人进行了现场试验，在 RFI 存在的情况下，自适应滤波方式能够极大改善脉冲星观测[9]。

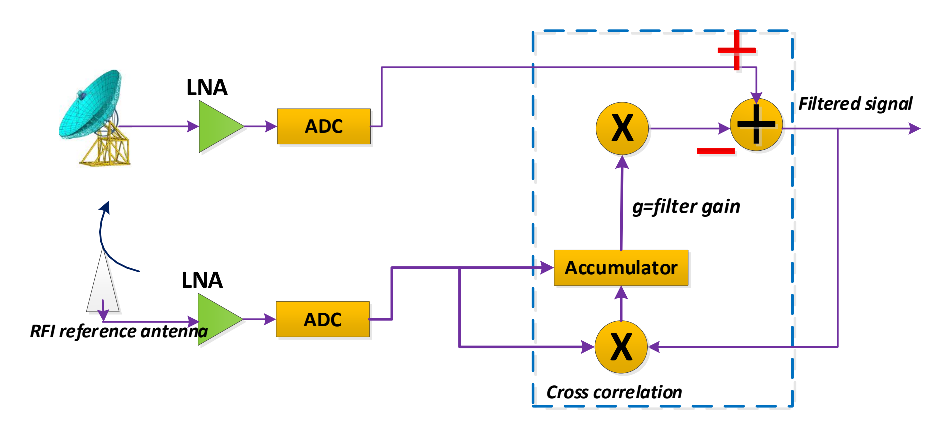


图 3 自适应滤波器基本设计图

Fig. 3 Adaptive filter basic design diagram

参考天线接收的信号可表示为：

$$R = R_{noise} + R_{rfi} + R_{ast} \#(1)$$

其中R表示参考天线接收的信号，R_{noise}表示参考天线的系统噪声，R_{rfi}表示参考天线接收的 RFI，R_{ast}表示参考天线接收到的天文信号，天文信号通过参考天线的旁瓣接收，其强度微弱可忽略不计。同理，主望远镜接收的信号可表示为：

$$M = M_{noise} + M_{rfi} + M_{ast} \#(2)$$

主望远镜和参考天线接收的系统噪声不相关，而接收的 RFI 具有相关性。为了消除主天线中的 RFI，使其对准 RFI 源，并通过不断优化参考天线的指向及极化方向可提高参考天线中 RFI 信噪比。

自适应滤波具体实现步骤:

(1) 在自适应相关环路中确定复增益系数g（该系数是不断迭代优化的）与参考天线接收的信号R相乘，逐步迭代$g \cdot R_{rfi}$，使其大小接近主天线的 RFI 信号，能最大程度消除主天线某一特定方向的 RFI。

(2) 当g达到最优值时，$g \cdot R_{rfi}$基本接近M_{rfi}，天文信号与参考信号的差值即为去除特定干扰后的天文信号。

(3) 图 3 为自适应滤波器基本设计图，将原始参考信号与滤波器输出相关，当相关系数为零时，达到最佳增益。

相关项计算如公式（3）所示，

$$C(g) = (M - g \cdot R) \cdot R \#(3)$$

当$C(g) = 0$时，可剔除主天线的 RFI，即

$$0 = M - g \cdot R \Rightarrow M = g \cdot R \Rightarrow g = \frac{M}{R} \#(4)$$

代入天线接收的各个分量得：

$$g = \frac{M_{noise} + M_{rfi} + M_{astr}}{R_{noise} + R_{rfi} + R_{astr}} = \frac{(M_{noise} + M_{rfi} + M_{astr}) \cdot (R_{noise} + R_{rfi} + R_{astr})}{(R_{noise} + R_{rfi} + R_{astr}) \cdot (R_{noise} + R_{rfi} + R_{astr})} \#(5)$$

因为两个天线的系统噪声、天文信号、其他信号均与 RFI 不相关，参考天线中的天文信号非常微弱，可以忽略不计，即$R_{ast} = 0$，那么上式进一步简化得：

$$g = \frac{M_{rfi} \cdot R_{rfi}}{R_{noise}^2 + R_{rfi}^2} \#(6)$$

令$M_{rfi} = \xi R_{rfi}$， g可以写为：

$$\mathrm{g} = \frac{\xi}{1 + 1/INR} \#(7)$$

其中 INR(Interference-to-Noise Ratio)是参考天线中的干扰噪声比，等于R_{rfi}^2/R_{noise}^2，可以明显看出，当 INR 越大，g值越小，自适应滤波器效果越好。输出信号中 RFI 的残留量为

$$\begin{aligned} S_{rfi} &= M_{rfi} - g \cdot R_{rfi} \\ &= M_{rfi} - \frac{\xi}{1 + \frac{1}{INR}} \cdot R_{rfi} \\ &= \frac{M_{rfi} + M_{rfi} \cdot INR - \xi INR \cdot R_{rfi}}{1 + INR} \\ &= \frac{M_{rfi} + M_{rfi} \cdot INR - \frac{M_{rfi}}{R_{rfi}} \cdot INR \cdot R_{rfi}}{1 + INR} \end{aligned}$$

$$= \frac{M_{rfi}}{1 + INR} \#(8)$$

在天文观测中，主天线指向待观测的天体，M_{rfi}在观测期间可以认为是恒定量，S_{rfi}与INR成反比，即INR越大，S_{rfi}越小，说明提高RFI在参考天线中信噪比能提高主天线RFI抑制能力。

影响自适应滤波性能的几个因素：

（1）混合 RFI 源

自适应滤波方法消除单源RFI效果较好，但是面对复杂的RFI环境，其抑制RFI的性能会大幅度下降。对于混合 RFI，当两个干扰信号处于相同频域时，振幅大的干扰将被消减。对于天文台址附近 RFI 来源复杂，且与天文信号相比具有较高的功率，在此类条件下自适应滤波方法RFI抑制效果一般[10]。

（2）多路径传播影响

多径传播是指由于信号到达接收天线的路径不同，而存在同一信号的多个副本的现象，通常是由大型陆地物体 (山脉、建筑物)、大型水体 (湖泊、泻湖) 和电离层反射产生的。

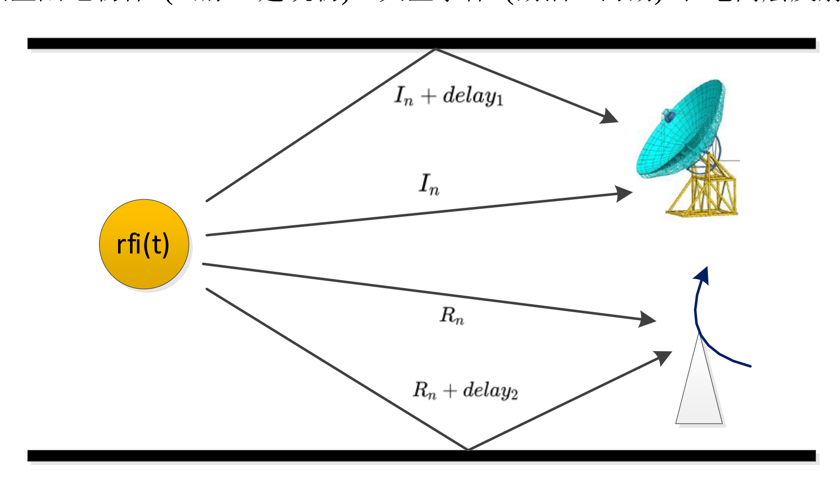


图 4 RFI 多径传播情况

Fig.4 RFI multipath propagation situation

图4描述了多路径干扰信号传播场景。因为到达望远镜和参考天线的多个副本有不同的时间延迟。如果延迟足够大，那么添加的信号就可以被望远镜视为完全不相关。在这种情况下，多路径情况相当于多个 RFI 源，RFI 信号种类繁多，难以获得滤波器多径传播的通用表达式。防止多径传播的最佳措施是将望远镜和参考天线放置在尽可能近的位置，以减小传播距离差异，并通过增加 FFT 点数来提高系统的频率分辨率，或在FFT点数不变的条件下减少系统带宽[10]。

（3）违反线性时不变(LTI)条件

假设干扰传播路径至少在滤波器收敛的时间内满足 LTI 条件。如果不满足 LTI 条件，

则表示 RFI 路径在某些地方呈现非线性或时变的传播特性。然而，没有任何自然介质表现出足够强的非线性，或具有快速变化的传播特性，而影响滤波器的性能，特别是考虑到最终算法收敛时间是毫秒级的。当望远镜在跟踪天体移动时，捕获 RFI 的旁瓣随望远镜一起移动，被测 RFI 的路径特性会随时间逐渐变化。自适应滤波方法需具备相应计算能力，以使其不会因望远镜的缓慢移动而受到明显影响。

接收机前端的一些模拟元件是信号饱和的，如放大器和频率混频器。当一个分量达到饱和时，输入信号的幅值就会被削去，从而产生谐波失真，将会影响系统的 LTI 条件。这种失真被视为附加的损坏信号，其频率是原始输入信号中心频率的几倍。此外，由于数字化后的混叠等原因，谐波信号往往会被混入基带从而破坏天文数据。主信道 RFI 谐波失真会严重破坏滤波器性能，附加的谐波信号不会出现在参考信道中，使得这类 RFI 更难被抑制[10]。

3.2 空间滤波

空间滤波方法主要针对多通道射电天文观测中的 RFI 抑制问题[11-13]，该技术适用于干涉射电望远镜阵列，如荷兰的 Westerbork 综合射电望远镜(WSRT)，美国的甚大阵列(VLA)，或配置了相控阵接收系统的单天线望远镜。图 5 显示了空间滤波技术在天线阵列系统中的应用。

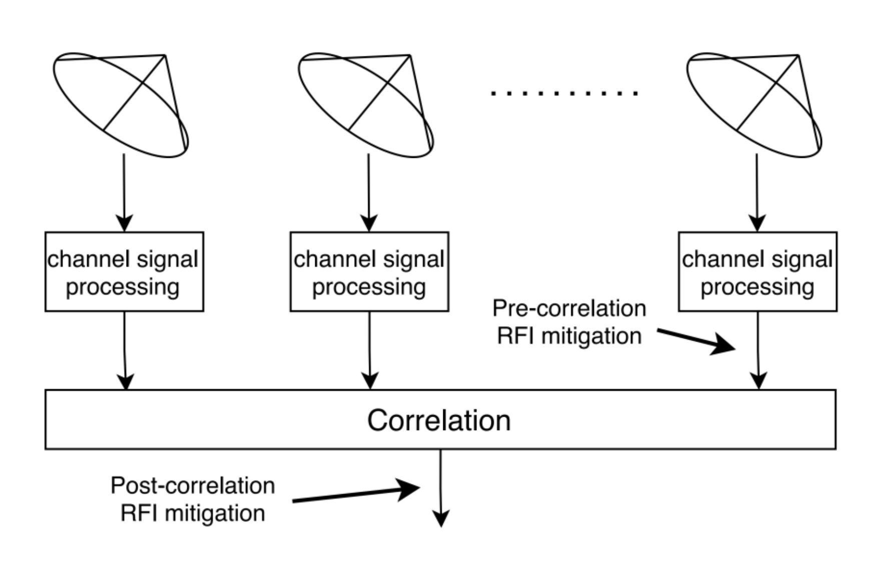


图 5 天线阵列系统在不同阶段抑制 RFI

Fig.5 Antenna array system suppresses RFI at different stages

射电望远镜接收的信号由三部分组成，射电源信号（拟观测目标）、系统噪声以及 RFI。系统噪声由宇宙背景噪声、大气噪声、接收机噪声等噪声分量组成。根据中心极限定理，系统噪声是时间独立的且高斯分布的信号，在短时间间隔上是短时平稳的，且所有的射电源都

被假定在统计学上彼此独立[12]。

射电天文学中的 RFI 特征是各种各样的。选择中心窄带高斯平稳、二阶非周期和循环平稳信号三种不同的信号特性来建模 RFI。

天线阵列共有M个天线，其中$x(t)=[x_1(t)\cdots x_M(t)]^T$表示的是$M\times 1$个天线阵列的输出，其中第$k$个天线在时间$t$时刻的输出为$x_k(t)$，那么$x(t)$可以被写为

$$\mathrm{x(t)=A_c\cdot c(t)+A_r\cdot r(t)+n(t)}\#(9)$$

其中

(1) $c(t)=\left[c_1(t)\cdots c_{N_c}(t)\right]^T$表示在$t$时刻时，射电源的特征向量，大小为$N_c\times 1$。

(2) $A_c=\left[c_1\left(t,\theta_{c_1},\phi_{c_1}\right)\cdots c_{N_c}\left(t,\theta_{c_{N_c}},\phi_{c_{N_c}}\right)\right]$是射电源的空间特征向量矩阵，大小为$M\times N_c$。

其中$c_n\left(t,\theta_{c_n},\phi_{c_n}\right)=\left[c_{n,1}\left(t,\theta_{c_n},\phi_{c_n}\right)\cdots c_{n,M}\left(t,\theta_{c_n},\phi_{c_n}\right)\right]^T$是第$n$个射电源的空间特征向量。同理，$A_r$和$r(t)$分别代表着 RFI 的空间特征向量矩阵以及在$t$时刻 RFI 的信号特征向量。

(3) $n(t)=[n_1(t)\cdots n_M(t)]^T$是系统噪声在$t$时刻的向量，大小为$M\times 1$。

通过协方差矩阵描述射电源在时间维度上的相关性，即

$$R(t,\tau)=E\{x\left(t+\frac{\tau}{2}\right)\cdot x^H\left(t-\frac{\tau}{2}\right)\}\#(10)$$

其中$R(t,\tau)$描述在t时刻τ范围内射电源信号的相关性，x^H代表x的共轭转置。假设 RFI、射电源和系统噪声三者相互独立，即互不相关，那么

$$E\{c_n\left(t+\frac{\tau}{2}\right)r_m^*\left(t-\frac{\tau}{2}\right)\}=0,\quad \forall\tau,n\in[1\ldots N_c],m\in[1\ldots N_r]$$
$$E\{c_n\left(t+\frac{\tau}{2}\right)n_m^*\left(t-\frac{\tau}{2}\right)\}=0,\quad \forall\tau,n\in[1\ldots N_c],m\in[1\ldots M]$$
$$E\{r_n\left(t+\frac{\tau}{2}\right)n_m^*\left(t-\frac{\tau}{2}\right)\}=0,\quad \forall\tau,n\in[1\ldots N_r],m\in[1\ldots M]\#(11)$$

从线性代数的观点来看，相控天线阵列射电望远镜数据协方差矩阵可以看作是线性变换矩阵。这个矩阵生成一个数据向量空间，该向量空间由 RFI 子空间、射电源子空间和系统噪声子空间组成。图 5 展示了一个无噪声场景下的二维数据向量空间的示例。红色和黑色矢量分别代表与射电源和射频干扰源相关的向量。

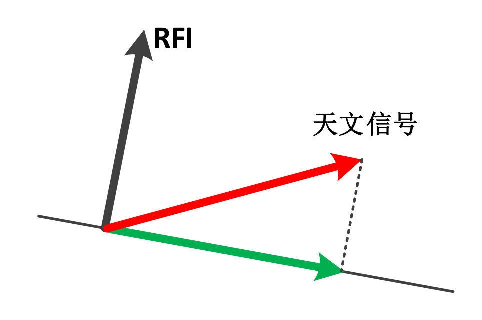


图 6 二维数据向量空间的正交投影

Fig.6 Orthogonal projection of a two-dimensional data vector space

使用正交投影技术减少干扰，将数据向量空间投影到与 RFI 子空间正交的子空间上，投影的 RFI 子空间完全为零，由此产生的数据只包含射电源。

从图 6 上可以清楚地看出，在投影后，恢复的射电源能量取决于两个向量的夹角大小，即

$$P_{proj}^2 = P^2 \cdot sin\left(\theta_{a_c,a_{r_n}}\right)\#(12)$$

其中P^2代表着射电源的功率，$\theta_{a_c,a_{r_n}}$代表着 RFI 向量与射电源向量的夹角大小，在$0\sim90^o$范围内，$\theta_{a_c,a_{r_n}}$越大，最终投影得到的射电源功率越大。但是射电源子空间和干扰子空间之间的夹角难以测量，需要借助投影矩阵间接计算。

设投影矩阵为 P，它的特征值是 0 或 1。特征值为 0 对应的特征向量生成投影核子空间，特征值为 1 对应的特征向量生成值域子空间。因此投影矩阵的秩等于它的值域子空间的维数。如果 H 是投影范围子空间基 ，K 是投影核子空间基，则有：

$$\mathrm{P}\cdot\mathrm{H}=\mathrm{H}$$

$$\mathrm{P}\cdot\mathrm{K}=0\#(13)$$

由基矩阵H生成的投影矩阵定义为

$$P = H(H^TH)^{-1}H^T\#(14)$$

其值域是正交于H的子空间的正交投影为$P^{\perp}$，定义为:

$$P^{\perp} = I - H(H^TH)^{-1}H^T\#(15)$$

在多干扰的情况下，干扰子空间是多维的，这个子空间是由存储在矩阵A_r中的独立 RFI 向量集合生成的。因此，投影矩阵可表示为

$$P_r^{\perp} = I - A_r(A_r^TA_r)^{-1}A_r^T\#(16)$$

在预相关阶段，可用数据为天线阵列输出数据向量$x(t)$，将其投影到 RFI 子空间的正交空间上，修正后的数据向量$x_{clean}(t)$为

$$x_{clean}(t) = P_r^{\perp}x(t)\#(17)$$

从以上内容可知，空间滤波需要预先知道干扰源来源，以便确定投影矩阵，对于复杂的干扰环境，干扰源具体位置测定是非常困难的。

4 RFI 后相关阶段

由于 RFI 信号强度通常比天文信号大几个数量级，阈值法是一种有效抑制 RFI 的方法。阈值水平通常是根据统计量决定的，可以用一段数据的均值、均方根等作为阈值参考值，超出这个范围之外的值都标记为 RFI。阈值法不足之处是会误将天文信号标记为 RFI，在后相关阶段降低 RFI 检测的错误率是研究 RFI 抑制问题的一个重要方向。

4.1 VarThreshold 算法

在频域和时域上，RFI 数据往往会影响相邻的数据，因为其强度比较低，往往不会被标记为 RFI，增加了检测 RFI 的错误率。VarThreshold 是一种组合阈值算法，其思想是当组合样本的某一统计属性超过某个限制时，将标记样本的组合为 RFI[14,15]。 假设A和B是相邻的样本采样点，常规阈值法是分别查看样本A和B是否超过设定的阈值水平线，如果超过，则标记为 RFI。但对于组合阈值来说，只有样本A和B都超过阈值，才可以将这一组合的样本都标记为 RFI。如果样本A和B没有被标记为 RFI，它们将和下一个相邻的采样点C进行组合，继续与阈值进行对比，迭代多次完成 RFI 标记，算法如公式(18)所示。

$$flag\ \upsilon_M(\upsilon,t) = \exists i \in \{0,\dots M-1\}: \forall j \in \{0,\dots M-1\}: \left|R\left(\upsilon + a\vartriangle\upsilon, t\right)\right| > \chi_M \#(18)$$

上式表达的是从t时刻起，M个数的范围内，如果该范围内所有采样点的绝对值大于χ_M，那么将这M个数都标记为 RFI。

VarThreshold 算法用一组严格递减的阈值序列$\chi_i(i=1,\dots,N)$判定某个采样点是否应被标记为 RFI，如公式（19）所示。

$$\chi_i = \frac{\chi_1}{\rho^{\log_2 i}} \#(19)$$

其中χ_1是单采样点阈值，当$\rho = 1.5$时效果最好，χ_1的取值是根据数据的统计水平量决定，即选取的阈值使 RFI 的假阳率最小。

4.2 SumThreshold 算法

SumThreshold 算法也是一种组合阈值算法[14,15]，与 VarThreshold 算法求解阈值序列类

似，只是ρ值不同，当$\rho = 1.2$时效果最好，求出一组阈值$T = \{T_1, T_2, T_4, \ldots\}$，通过迭代的方式与样本进行比较，即$T_1$与单样本比较，$T_2$与两个样本的均值比较，$T_m$与$m$个样本的均值比较，当样本均值大于阈值时，被标记为 RFI，且不参与后续的迭代计算过程。以一维序列为例，设序列为{1,7,4,3,2,1,1,6,1,1,3,2,3,2}，阈值χ为$\{\chi_1 = 6, \chi_2 = 3, \chi_4 = 2\}$。

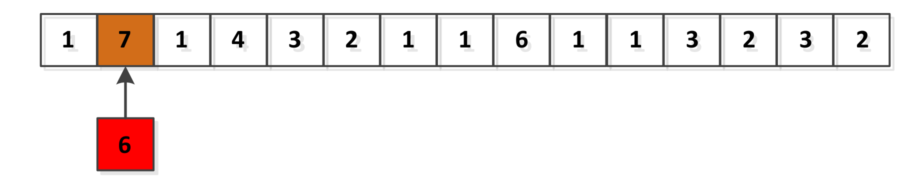


图 7 第一轮，阈值为$\chi_1 = 6$，只有 7 被标记

Fig.7 In the first round, the threshold is $\chi_1 = 6$ and only 7 is marked

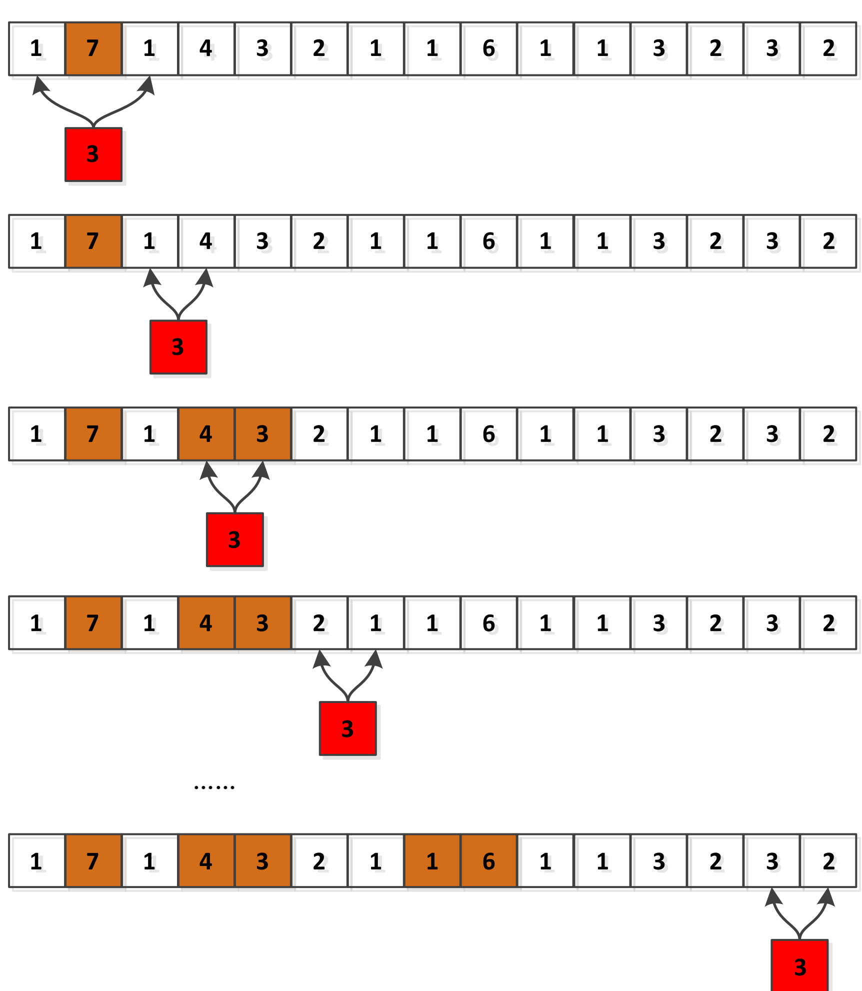


图 8 第二轮，阈值为$\chi_2 = 3$

Fig.8 In the second round, the threshold value is $\chi_2 = 3$

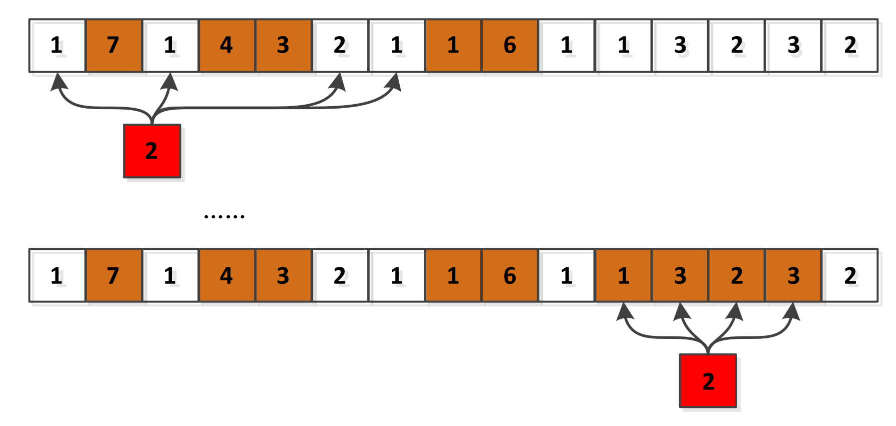


图 9 第三轮，阈值为$\chi_4 = 2$

Fig.9 In the third round, the threshold value is $\chi_4 = 2$

图 7，图 8 和图 9 展示了 SumThreshold 算法的过程。从算法角度考虑，该算法的时间复杂度为$O(n^2)$。为了减少计算速度，当阈值M取 1,2,4,8, 16,32,64,128,256…时，时间复杂度降至$O(nlogn)$。

4.3 奇异值分解

奇异值分解（Singular Value Decomposition，SVD）是线性代数中一种重要的矩阵分解，能够提取信息，简约数据，去除噪声等，可将数据映射到低维空间[16]。SVD 可分解为

$$\boldsymbol{A} = \boldsymbol{U\Sigma V^T} \#(20)$$

如公式(20)，SVD 将原始矩阵$\boldsymbol{A}$分解为三个矩阵：$\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V^T}$，其中$\boldsymbol{U}, \boldsymbol{V}$是正交矩阵，$\boldsymbol{\Sigma}$是对角矩阵，对角线上的元素为奇异值。绝大部分的 RFI 信号在时域或频域上的振幅较大。用$\boldsymbol{A}$代表接受的天文信号，将其进行奇异值分解，即$\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$，将$\boldsymbol{\Sigma}$中最大特征值置为 0，将其逆向恢复后，得到新的矩阵$\hat{A}$，那么原矩阵$\boldsymbol{A}$中幅值较大的信号（RFI）会被抑制。

当 RFI 信号足够强，且$\sigma_{RFI} \cdot \sigma_{astro} = 0$时，该算法效果最好，但该方法仅适用于宽带 RFI 信号，对于遵从高斯分布的 RFI 信号并不适用[14,15]。

5 基于机器学习的 RFI 抑制方法

随着机器学习和深度学习等技术快速发展以及天文数据的急速增长，两者的结合已成必然趋势，通过对大量的数据进行学习和训练后，能有效地标记天文信号中的 RFI，并进行特征提取。

5.1 主成分分析法

主成分分析（Principal Component Analysis，PCA）是一种用于数据特征降维的方法，常用于减少数据集的维数，同时保持数据集对方差贡献最大的特征。这样低阶成分往往能够保留数据的重要信息，方便特征提取和分类[17,18]。

文献[19]对生活中常见的九种瞬时 RFI 来源进行分类测试实验，在时域范围内分别使用标准 PCA 和核 PCA 进行测试。之后用一种聚类分离的方法比较了两种方法的聚类精度，如图 10 和图 11，发现核 PCA 在区分来源方面比标准 PCA 更好，能有效区分瞬时 RFI 来源。

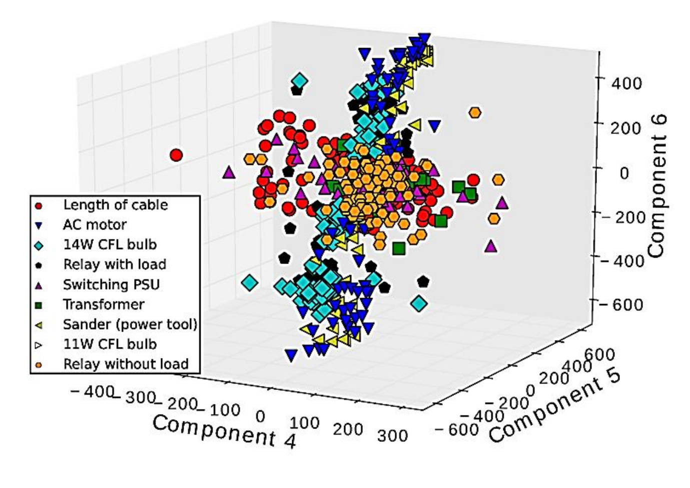


图 10 标准 PCA（来源文献[18]）

Fig. 10 Standard PCA

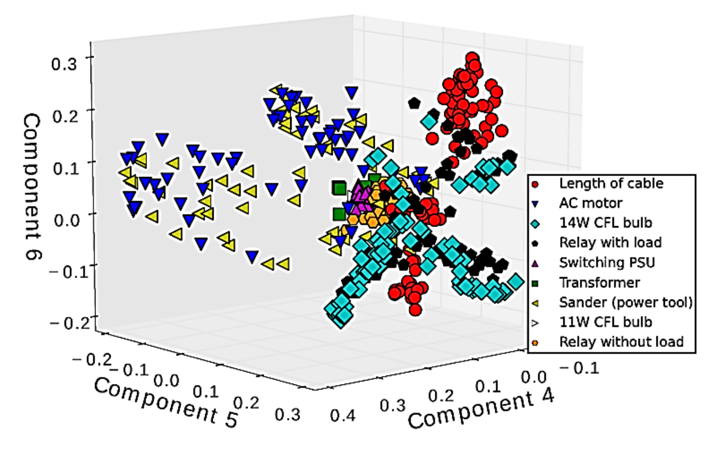


图 11 核 PCA（来源文献[18]）

Fig. 11 Kernel PCA

5.2 支持向量机

在机器学习中，支持向量机（Support Vector Machine，SVM）是在分类与回归分析中分析数据的监督式学习模型与相关的学习算法[20]。

文献[21]选取时间长度为300μs非 RFI 和 RFI 信号训练，提取均方根、均值、方差、均值与方差之比、偏度、峰度、最大振幅、最小振幅和峰值个数等信息，导入 SVM 训练模型，之后对测试数据进行 RFI 和非 RFI 分类，流程如图 12 所示。结果表明即使在非常低 INR 和 RFI 占空比小的情况下，该方法仍能准确地检测出 RFI，效果良好。

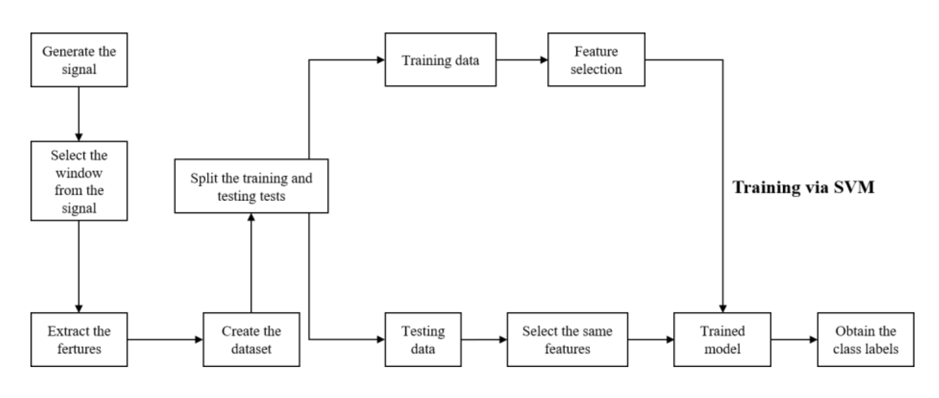


图 12 基于 SVM 的 RFI 抑制过程

Fig. 12 RFI suppression process based on SVM

5.3 神经网络

神经网络的基本结构包含输入层、隐藏层和输出层，由大量的节点和之间相互连接构成，是一种非线性统计性数据建模工具，常用来对输入和输出间复杂的关系进行建模，或用来探索数据的模式[22]。

文献[23]提出了全卷积神经网络(Deep Fully Convolutional Neural Network，DFCN，用于图像的分割，解决的核心问题是图像像素级别的分类。文献[24]利用 DFCN，如图 13，H 和 W 分别对应输入的时间和频率可见性维数，F为滤波器层数，L对应输入和全卷积层之间的层数总和，以振幅和相位为特征，瀑布图进行特征提取，并标记出 RFI。

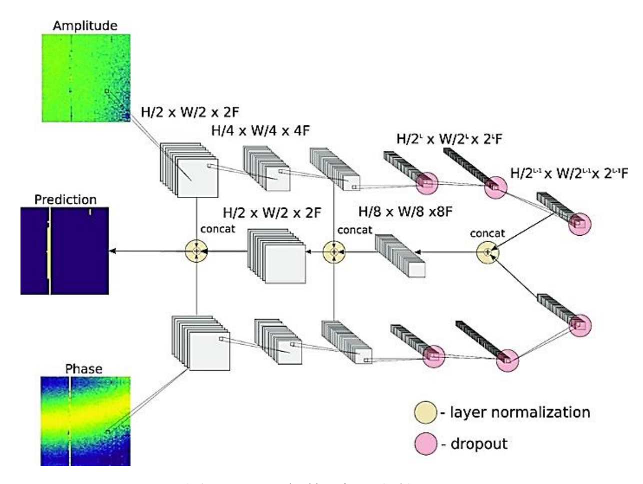


图 13 DFCN 架构(来源文献[23])

Fig.13 DFCN architecture

文献[25]提出的 U-Net 神经网络，是在卷积神经网络的基础上增加了上卷积的扩展路径，可以应对小样本的数据集进行较快、有效地图像分割，被用于医学细胞图像分割。文献[3]成功地应用了 U-Net 深度神经网络模型来识别单天线射电望远镜数据中的 RFI，其结构如图 14。

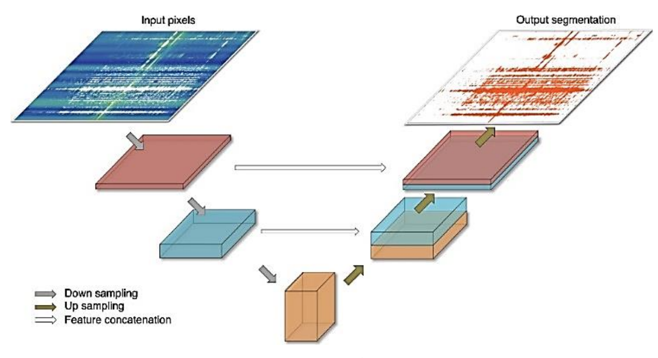


图 14 U-Net 架构(来源文献[3])

Fig.14 U-Net architecture

文献[26]发现使用卷积神经网络(Convolutional Neural Networks, CNN)在对 FAST 数据进行标记 RFI 时，常常标记错误，需要进一步人工检查，带来了显著的额外工作量，为了克服这一问题，提出了一种称为 RFI-Net 的神经网络模型如图 15 所示,结果表明，RFI-Net 模型要优于 U-Net，KNN(K-nearest Neighbour)和 SumThreshold 算法。

循环神经网络(Recurrent Neural Network，RNN)是一类用于处理序列数据的神经网络，对处理上下文关联的语音[27]数据有很好效果。天文信号也是一种时序数据，非常适和使用 RNN 对其进行处理，Burd 等人利用 RNN 对射电望远镜干涉阵列数据进行 RFI 检测，根据巨米波射电望远镜（Giant Metrewave Radio Telescope，GMRT）的 610MHz 的 RFI 振幅信息，区分 RFI 和非 RFI 数据[28]。

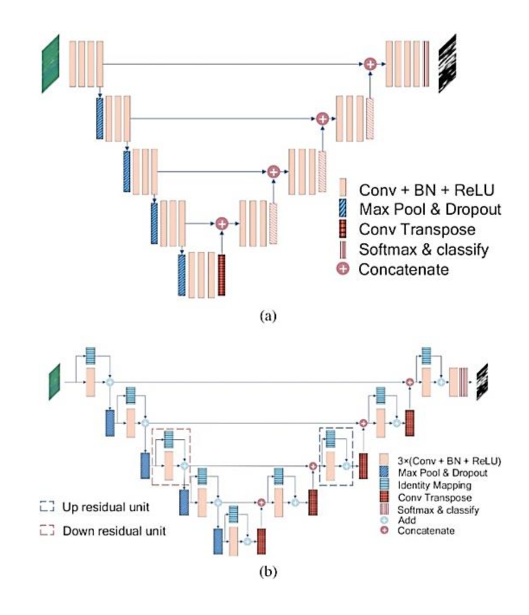


图 15 RFI-Net 架构(来源文献[25])

Fig. 15 RFI-Net architecture

6 总结

本文系统阐述了射电天文中抑制和标记 RFI 算法和技术，分析了在主动预防阶段、预相关阶段、后相关阶段 RFI 抑制方法的优势和不足。采用有效地屏蔽手段可以在主动预防阶段阻止大部分 RFI 进入系统；预相关阶段基于各个天线中的天文信号具有相关特性抑制 RFI；在后相关阶段大部分算法基于阈值标记 RFI。通过机器学习相关技术对 RFI 进行标记和识别是目前的研究热点，基于海量天文数据进行训练能够极大提高标记 RFI 的正确率，但不足是前期手动标记训练样本中 RFI 和天文信号十分耗费时间，目前可用的训练样本有限。

RFI 抑制问题需要在不同阶段采用多种方法协同解决，从射电望远镜台址周边的无线电环境保护到最终天文数据处理过程各个环节都需要考虑 RFI 问题。各天文台站需要结合实际需求和所处环境的电磁干扰情况，合理选择 RFI 抑制方法。

致谢： 感谢国家重点研发计划（2021YFC2203502）；国家自然科学基金（12173077,11873082,11803080,12003062）；新疆维吾尔自治区天山创新团队计划（2022D14020）；中国科学院青年创新促进会；国家重点研发计划(2018YFA0404704)；国家天文科学数据中心；中国科学院天文台站设备更新及重大仪器设备运行专项经费资助。本论文得到中国虚拟天文台、国家天文科学数据中心、中科院科学数据中心体系提供的数据资源和技术支持。

参考文献

[1] Wang Yu ,Zhang Hai-Yan , Hu Hao et al. Satellite RFI mitigation on FAST[J]. Research in Astronomy and Astrophysics,2021(21):18.

[2] Porko, Jukka-Pekka Göran.Radio frequency interference in radio astronomy[D], Aalto University, 2011:77

[3] J. Akeret ,C. Chang , A. Lucchi , A. Refregie. Radio frequency interference mitigation using deep convolutional neural networks[J]. Astronomy and Computing,2017(18):35-39.

[4] 胡浩,张海燕,黄仕杰.FAST 电波环境保护措施[J].深空探测学报,2020,7(02):152-157.

HU Hao， ZHANG Haiyan， HUANG Shijie.Protection Measures of FAST Radio Environment[J]. Journal of Deep Space Exploration,2020,7(02):152-157.

[5] 王娜.新疆奇台 110 米射电望远镜[J].中国科学:物理学 力学 天学, 2014, 44(08): 783-794.

WANG Na. Xinjiang Qitai 110 m radio telescope[J]. SCIENTIA SINICA Physica, Mechanica & Astronomica, 2014, 44(08): 783-794.

[6] 刘晔. QTT 台站限制区域电磁干扰影响分析[C].中国天文学会 2018 年学术年会.

LIU Ye. Analysis of electromagnetic interference impact in the restricted area of QTT stations[C]. Chinese Astronomical Society 2018 Annual Academic Conference.

[7]刘晔,刘奇.大口径射电望远镜台址电磁干扰预测方法[J].中国科学:物理学 力学 天文学,2019,49(09):121-129.

LIU Ye, LIU Qi. A prediction method for electromagnetic interference of large aperture radio telescope. [J]. SCIENTIA SINICA Physica, Mechanica & Astronomica,2019,49(09):121-129.

[8] Cecilia Barnbaum and Richard F. Bradley. A New Approach to Interference Excision in Radio Astronomy: Real-Time Adaptive Cancellation [J].The Astronomical Journal,1998(116):2598-2614.

[9] Kesteven, M. and Hobbs, G. and Clement, R. and Dawson, B. and Manchester, R. and Uppal, T. Adaptive filters revisited: Radio frequency interference mitigation in pulsar observations[J]. Radio Science,2005(40).

[10] Curotto Molina, Franco Andreas. Design, implementation and characterization of a radio frecuency interference digital adaptive filter using a field-programmable gate array[D]. Universidad de Chile,2019.

[11] Raza J, Boonstra A J, Van der Veen A J. Spatial filtering of RF interference in radio astronomy[J]. IEEE Signal Processing Letters, 2002, 9(2): 64-67.

[12] Hellbourg G. Radio Frequency Interference spatial processing for modern radio telescopes[D]. Universit éd'Orl éans, 2014.

[13] Kocz J, Briggs F, Reynolds J. Spatial filtering using a multibeam receiver[C]//Proceedings of RFI Mitigation Workshop PoS (RFI2010) proceedings. 2010, 1: 32.

[14] Offringa, A. R. , de Bruyn, A. G. , Biehl, M. et al. Post-correlation radio frequency interference classification methods[J]. Monthly Notices of the Royal Astronomical Society,2010(405):155-167.

[15] Offringa A R. Algorithms for radio interference detection and removal[D]. University of Groningen, 2012.

[16] Guo Q, Zhang C, Zhang Y, et al. An efficient SVD-based method for image denoising[J]. IEEE transactions on Circuits and Systems for Video Technology, 2015, 26(5): 868-880.

[17] Abdi H, Williams L J. Principal component analysis[J]. Wiley interdisciplinary reviews: computational statistics, 2010, 2(4): 433-459.

[18] Kherif F, Latypova A. Principal component analysis[M]//Machine Learning. Academic Press, 2020: 209-225.

[19] Czech Daniel, Mishra Amit Kumar, Inggs Michael. Time domain classification of transient radio frequency interference[C]. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS),2016:302-305.

[20] Pisner D A, Schnyer D M. Support vector machine[M]//Machine Learning. Academic Press, 2020: 101-121.

[21] Nazar I M, Aksoy M. Radio Frequency Interference Detection in Microwave Radiometry Using Support Vector Machines[J]. URSI GASS, 2020.

[22] Aggarwal C C. Neural networks and deep learning[J]. Springer, 2018, 10: 978-3.

[23] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[24] Kerrigan J, Plante P L, Kohn S, et al. Optimizing sparse RFI prediction using deep learning[J]. Monthly Notices of the Royal Astronomical Society, 2019, 488(2): 2605-2615.

[25] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[26] Yang Z, Yu C, Xiao J, et al. Deep residual detection of radio frequency interference for FAST[J]. Monthly Notices of the Royal Astronomical Society, 2020, 492(1): 1421-1431.

[27] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.2329, 2014.

[28] Burd P R, Mannheim K, März T, et al. Detecting radio frequency interference in radio‐antenna arrays with the recurrent neural network algorithm[J]. Astronomische Nachrichten, 2018, 339(5): 358-362.

Application of RFI Mitigation Technology in Radio Astronomy

Zhang Hailong[1,2,3,4], Zhang Yazhou[1,2], Wang Jie[1,4], Ye Xinchen[1,4], Wang Wanqiong[1], Li Jia[1], Zhang Meng[1,2], Du Xu[1,2]

(1 Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011,China; 2 University of Chinese Academy of

Sciences, Beijing 100049, China; 3 Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008,China;4 National Astronomical Data Center, Beijing 100101, China;)

Abstract: Aiming at the problem of radio frequency interference (RFI) in the process of radio astronomy observation, the RFI mitigation strategies of domestic and foreign stations are analyzed in detail. According to the RFI problems encountered in the actual observation process of each observatory, the prevention strategies and mitigation methods of RFI were studied from the aspects of active prevention stage, pre-correlation stage, post-correlation stage, machine learning and deep learning. The methods that can be adopted in the active prevention stage, adaptive filtering in the pre-correlation stage, spatial filtering method, VarThreshold, SumThreshold and singular value decomposition in the post correlation stage are systematically analyzed. The applications of principal component analysis based on machine learning, support vector machine, full convolution neural network, convolution neural network, u-net and other related technologies and methods in RFI signal processing are discussed.

Key words: RFI; filtering; thresholding; machine learning.

Aiming at the problem of radio frequency interference (RFI) in the process of radio astronomy observation, the RFI mitigation strategies of domestic and foreign stations are analyzed in detail. According to the RFI problems encountered in the actual observation process of each observatory, the prevention strategies and mitagation methods of RFI were studied from the aspects of active prevention stage, pre-correlation stage, post-correlation stage, machine learning and deep learning. The methods that can be adopted in the active prevention stage, adaptive filtering in the pre-correlation stage, spatial filtering method, VarThreshold, SumThreshold and singular value decomposition in the post correlation stage are systematically analyzed. The applications of principal component analysis based on machine learning, support vector machine, full convolution neural network, convolution neural network, u-net and other related technologies and methods in RFI signal processing are discussed.